

基于我国各地区普通高等教育发展状况分析

吕黄铭

云南财经大学, 云南 昆明
Email: 595145778@qq.com

收稿日期: 2021年1月5日; 录用日期: 2021年2月12日; 发布日期: 2021年2月19日

摘要

我国经济水平快速发展, 但一些地区经济发展水平还是不平衡, 特别是教育水平存在明显的差异, 本文就30个省份的教育水平进行分析, 主要通过类平均法还有Ward方法对数据进行聚类, 首先需要找好数据, 建立好模型, 之后运用SPSS的聚类分析得出结论, 来探讨教育发展的情况, 通过分析发现我国的教育水平发展十分不均衡。因此, 国家还有政府应该采取措施改善教育发展的情况, 让每个孩子都有享受平等教育的机会。

关键词

教育水平, 聚类分析, 类平均法, Ward方法, 平等教育

Based on the Analysis of the Development Status of General Higher Education in Various Regions of Our Country

Huangming Lv

Yunnan University of Finance and Economics, Kunming Yunnan
Email: 595145778@qq.com

Received: Jan. 5th, 2021; accepted: Feb. 12th, 2021; published: Feb. 19th, 2021

Abstract

My country's economic level is developing rapidly, but the level of economic development in some regions is still unbalanced, and especially there are obvious differences in the level of education. This article analyzes the education level of 30 provinces, mainly using the class average method and the Ward method to cluster the data. First, we need to find good data, establish a good model,

and then use the cluster analysis of SPSS to draw a conclusion to discuss the development of education. Through the analysis, it is found that the level of education in my country is developing very unevenly. Therefore, the country and the government should take measures to improve the development of education so that every child has an equal opportunity to enjoy education.

Keywords

Education Level, Cluster Analysis, Class Average Method, Ward Method, Equal Education

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

教育公平是整个社会还有人类赖以发展的基础，只有教育平等，人权才平等，思想境界才能提高，国家才能有文化核心竞争力，才能够创新，屹立于世界。只有坚持教育，让教育平等，才能加速全面建成小康社会。因此，我们必须对不同省份的教育水平做一个分析了解，让中央教育局知道需要制定怎样的教育政策，提供怎样的教育资源，可以帮助同学们快速发展，本文运用聚类分析进行聚类，得出哪些城市的教育水平比较好，哪些城市还需要加强，进而促进教育平等，社会和谐有序地发展。

2. 现状分析

在现在这样一个卧虎藏龙的时代，我认识到一些很多接受教育资源比我好的同学，的确他们在社会这场竞争中就是处于一个优势的状态，你不得不感叹他们的学识还有气质就是比你优秀。其实这归根于区域经济发展不均衡带来的负面影响，作为一个 13 亿人口的中国大国，在贫富差距如此悬殊，区域经济发展、教育资源分配如此不均衡的情况下，如果不着手从根本上解决区域发展不均衡，教育水平不平等，任由发展下去是会阻碍社会的健康发展，不利于国家的进步。如果经济发展成果不能由全体人民共享，而是集中在少数人少数地区的人民垄断，在经济全球化过程中利益受损的人很可能会做出一些无理还有极端的行为表达自己的不满，近年来欧洲右翼政党纷纷登上政治舞台、“英国脱欧”、“特朗普现象”、“贸易保护主义”等频频出现的根源就在于经济全球化的发展成果被少部分地区还有少部分的集团垄断，“红脖子们”未能共享经济发展的成果，偏远山区还有省份未能享受到平等的教育环境。其实，每一个孩子都应该拥有接受同能优质教育的权力，每一个省份的经济发展，教育的硬件措施都应该平等对待，这也是人类几千年来一直追求的境界“大道之行，天下为公”，而不是像一些西方国家一样用所谓的“快乐教育”和学费高昂的“精英教育”通过种种看得见的看不见的制度和套路实现安排“龙生龙，凤生凤”的阶级固化，只有从解放人民的思想，发展文化，社会才能实现大同。

3. 研究目的和意义

受教育的机会应该是平等的，不应该出现由于经济的差异出现不平等的现象。因此需要迫切了解每个省的教育水平，教育资源情况，国家根据相应的分析制定合理的政策，帮助一些省份地区教育水平的发展、提高。只有坚持教育平等，赋予每个受教育者同等教育的机会，才能促进社会成员的潜能、积极性和创造力，才能提高全员的素质，也不会说出现“老人摔倒该不该扶”，“豆腐渣工程”、“江歌案”等道德缺失，教育缺失的事件。教育兴则国兴，教育强则国强，习近平主席也一直强调教育发展的战略

核心地位，教育上来了，自然国家水平就上来了，人民生活水平提高，进入全体富裕，加快全面进入到社会主义现代化社会。

4. 综合评测指标的确定

指标的数据选自《中国统计年鉴》和《中国教育统计年鉴》。其中 x_1 为每百万人口的高等院校数； x_2 为每 10 万人口高等院校毕业生数； x_3 为每 10 万人口高等院校招生数； x_4 为每 10 万人口高等院校在校生数； x_5 为每 10 万人口高等院校教职工数； x_6 为每 10 万人口高等院校专职教师数； x_7 为高级职称占专职教师的比例； x_8 为平均每所高等院校的在校生数； x_9 为国家财政预算内普通高等经费占国内生产总值的比例； x_{10} 为生均教育经费[1]。

由图 1，我们先对指标运用因子分析，得出相关系数矩阵，然后根据相关性选择有很强相关性的变量进行聚类，运用类平均法还有 Ward 方法进行聚类，然后根据聚类结果进行分析得出结论。

地区	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
北京	5.96	310	461	1557	931	319	44.36	2615	2.2	13631
上海	3.39	234	308	1035	498	161	35.02	3052	0.9	12665
天津	2.35	157	229	713	295	109	38.4	3031	0.86	9385
陕西	1.35	81	111	364	150	58	30.45	2699	1.22	7881
辽宁	1.5	88	128	421	144	58	34.3	2808	0.54	7733
吉林	1.67	86	120	370	153	58	33.53	2215	0.76	7480
黑龙江	1.17	63	93	296	117	44	35.22	2528	0.58	8570
湖北	1.05	67	92	297	115	43	32.89	2835	0.66	7262
江苏	0.95	64	94	287	102	39	31.54	3008	0.39	7786
广东	0.69	39	71	205	61	24	34.5	2988	0.37	11355
四川	0.56	40	57	177	61	23	32.62	3149	0.55	7693
山东	0.57	58	64	181	57	22	32.95	3202	0.28	6805
甘肃	0.71	42	62	190	66	26	28.14	2657	0.73	7282
湖南	0.74	42	61	194	61	24	33.06	2618	0.47	6477
浙江	0.86	42	71	204	66	26	29.94	2363	0.25	7704
新疆	1.29	47	73	265	114	46	25.93	2060	0.37	5719
福建	1.04	53	71	218	63	26	29.01	2099	0.29	7106
山西	0.85	53	65	218	76	30	25.63	2555	0.43	5580
河北	0.81	43	66	188	61	23	29.82	2313	0.31	5704
安徽	0.59	35	47	146	46	20	32.83	2488	0.33	5628
云南	0.66	36	40	130	44	19	28.55	1974	0.48	9106
江西	0.77	43	63	194	67	23	28.81	2515	0.34	4085
湖南	0.7	33	51	165	47	18	27.34	2344	0.28	7928
内蒙古	0.84	43	48	171	65	29	27.65	2032	0.32	5581
西藏	1.69	26	45	137	75	33	12.1	810	1	14199
河南	0.55	32	46	130	44	17	28.41	2341	0.3	5714
广西	0.6	28	43	129	39	17	31.93	2146	0.24	5139
宁夏	1.39	48	62	208	77	34	22.7	1500	0.42	5177
贵州	0.64	23	32	93	37	16	28.12	1469	0.34	5415
青海	1.48	38	46	151	63	30	17.87	1024	0.38	7368

Figure 1. Data of general higher education development in different regions of China

图 1. 我国各地区普通高等教育发展状况数据

见图 2，我们可以清楚的看出变量之间都有很强的相关性，所以直接运用变量进行聚类分析，去分析出不同省份的教育水平的情况，并根据结论给出意见。

因子分析

		相关矩阵									
		x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
相关	x1	1.000	0.943	0.953	0.959	0.975	0.980	0.407	0.066	0.888	0.659
	x2	0.943	1.000	0.995	0.995	0.974	0.987	0.614	0.350	0.804	0.599
	x3	0.953	0.995	1.000	0.999	0.983	0.981	0.626	0.344	0.823	0.617
	x4	0.959	0.995	0.999	1.000	0.988	0.986	0.610	0.326	0.828	0.612
	x5	0.975	0.974	0.983	0.988	1.000	0.999	0.560	0.241	0.859	0.617
	x6	0.980	0.970	0.981	0.986	0.999	1.000	0.550	0.222	0.869	0.615
	x7	0.407	0.614	0.626	0.610	0.560	0.550	1.000	0.779	0.366	0.154
	x8	0.066	0.350	0.344	0.326	0.241	0.222	0.779	1.000	0.112	0.052
	x9	0.868	0.804	0.823	0.828	0.859	0.869	0.366	0.112	1.000	0.682
	x10	0.659	0.599	0.617	0.612	0.617	0.615	0.154	0.052	0.682	1.000

Figure 2. Correlation coefficient matrix
图 2. 相关系数矩阵

5. 聚类分析的基本思想还有方法

5.1. 聚类分析的定义

聚类分析是根据研究对象的特征对它进行分类的方法。聚类分析的目标就是在相似的基础上收集数据来分类。聚类源于很多领域，包括数学，计算机科学，统计学，生物学和经济学。在不同的应用领域，很多聚类技术都得到了发展，这些技术方法被用作描述数据，衡量不同数据源间的相似性，以及把数据源分类到不同的簇中[2]。

5.2. 聚类分析的步骤

- 1) 选择描述事物对象的变量指标；
- 2) 建立样品系列矩阵查看相关性；
- 3) 确定数据是否要进行标准化；
- 4) 确定表示对象距离或者选择相似程度比较高的统计量；
- 5) 选择适当的聚类方法，进行聚类；
- 6) 根据聚类结果给出相关结论。

5.3. 计算对象之间的距离

设 $x = (x_1, x_2, \dots, x_p)'$ 和 $y = (y_1, y_2, \dots, y_p)'$ 为两个样品，则所定义的距离一般应该满足如下三个条件：

- 1) 非负性： $d(x, y) \geq 0$ ， $d(x, y) = 0$ 当且仅当 $x = y$ ；
- 2) 对称性： $d(x, y) = d(y, x)$ ；
- 3) 三角不等式： $d(x, y) \leq d(x, z) + d(z, y)$ ；

在聚类过程中，相距较近的样品倾向于归于一类，相距比较远的归于一类。常用的距离如下几种表示：

5.3.1. 明考夫斯基距离

x 和 y 之间的明考夫斯基距离定义为：

$$d(x, y) = \left[\sum_{i=1}^p |x_i - y_i|^q \right]^{1/q}$$

这里 $q \geq 1$ 。明氏距离有以下三种特殊的形式：

当 $q=1$ 时， $d(x, y) = \sum_{i=1}^p |x_i - y_i|$ ，称为绝对值距离，常被形象地称作“城市街区”；

当 $q=2$ 时， $d(x, y) = \left[\sum_{i=1}^p |x_i - y_i|^2 \right]^{\frac{1}{2}} = \sqrt{(X - Y')(X - Y)}$ ，这是欧式距离，它是聚类分析中最常用的一种距离；

当 $q = \infty$ 时， $d(x, y) = \max |x_i - y_i|$ ，称为切比雪夫距离。

5.3.2. 兰氏距离

当所有的数据皆为正时，可以定义 x 与 y 之间的兰氏距离为

$$d(x, y) = \sum_{i=1}^p \frac{|x_i - y_i|}{x_i + y_i}$$

该距离与各变量的单位无关，且适用于高度偏斜或者含有异常值的数据。

5.3.3. 马氏距离

x 与 y 之间的马氏距离为

$d(x, y) = \sqrt{(X - Y')S^{-1}(X - Y)}$ ，其中 S 为样本协方差矩阵，使用马氏距离的好处是考虑到了各变量之间的相关性，并且与变量的单位无关。

5.4. 聚类的方法还有思想

5.4.1. 类平均法

类平均法有两种定义，一种定义方法是把类与类之间的距离定义为所有样品对之间的平均距离，定义为：

$$D_{kl} = \frac{1}{n_k n_l} \sum_{i \in G_k, j \in G_l} d_{ij}$$

其中 n_k 和 n_l 分别为类 G_k 和 G_l 的个数。

另外一种定义方法是定义类与类之间的平方距离为样品对之间平方距离的平均值，即是：

$$D_{kl}^2 = \frac{1}{n_k n_l} \sum_{i \in G_k, j \in G_l} d_{ij}^2$$

5.4.2. 离差平方和法

类中各样品到类重心的平方欧式距离之和称为离差平方和。设类 G_k 和 G_l 合并成新类 G_m ，则 G_k 和 G_l 和 G_m 的离差平方和分别是

$$W_k = \sum_{i \in G_k} (x_i - \bar{x}_k)' (x_i - \bar{x}_k)$$

$$W_l = \sum_{i \in G_l} (x_i - \bar{x}_l)' (x_i - \bar{x}_l)$$

$$W_m = \sum_{i \in G_m} (x_i - \bar{x}_m)' (x_i - \bar{x}_m)$$

如果 G_k 和 G_l 这两类相距较近，则合并后的离差平方和 $W_m - W_k - W_l$ 应当较小，否则，应当比较大。于是我们定义 G_k 和 G_l 之间的平方距离为：

$$D_{kl}^2 = W_m - W_k - W_l$$

这种系统聚类的方法是 Ward 方法，实际上这种聚类的效果会比较好。

思想：让 n 个样品各自成一类，然后每次缩小一个类，每缩小一类，离差平方和就要增大，选择使得 L 增加最小的两类合并，知道所有的样品归为一类为止[3]。

6. 结论

案例研究结果

1) 各地区高等教育发展状况存在较大的差异，高校资源的地区分布很不均衡，如果用类平均法把 30 个地区分为 5 类，其结果为：

第一类：北京、上海；第二类：天津、陕西、辽宁、吉林、黑龙江、湖北、江苏、四川、山东、甘肃、浙江、福建、云南、海南、青海；第三类：广东；第四类：其他地区；第五类：西藏。

如图 3 结果还有树状图可以看出，北京，上海的高等教育状况与其他地区相比有非常大的不同，主要集中在每百万人口的学校数量和每 10 万人口的学生数量以及国家财政预算内普通高校经费占国内生产总值的比例远远高于其他地区，这与北京上海是我国政治文化中心的情况是类似的，北京，上海的高校数目也是位列全国之首，第二类的天津、陕西、辽宁等城市作为第二类，仅次于北京上海，高等教育水平中上，有比较多的资源，比如现在在西安，武汉的高校数目也是比较大的，毕竟之前西安是古代的首都，也是文化的政治中心，第三类是广东，这个跟现在有一点出入，自从改革开放起来，广东的经济快速发展，现在已经成为了一线城市，第四类的地区资源比较匮乏，国家财政还有高校资源相对比较稀缺，是国家教育大力扶植的对象。对于西藏，作为一个非常特殊的民族地区，其高等教育情况和其他地区不同，被单独聚为一类，主要表现在资源比较匮乏，人口相对比较少，经费比较稀缺，高等院校的规模较小，师资力量薄弱，最近不是一直向西藏提供支教，来提升西藏的教育水平。

	群集成员	
案例		5 群集
1: 北京		1
2: 上海		1
3: 天津		2
4: 陕西		2
5: 辽宁		2
6: 吉林		2
7: 黑龙江		2
8: 湖北		2
9: 江苏		2
10: 广东		3
11: 四川		2
12: 山东		2
13: 甘肃		2
14: 湖南		4
15: 浙江		2
16: 新疆		4
17: 福建		2
18: 山西		4
19: 河北		4
20: 安徽		4
21: 云南		2
22: 江西		4
23: 海南		2
24: 内蒙古		4
25: 西藏		5
26: 河南		4
27: 广西		4
28: 宁夏		4
29: 贵州		4
30: 青海		2

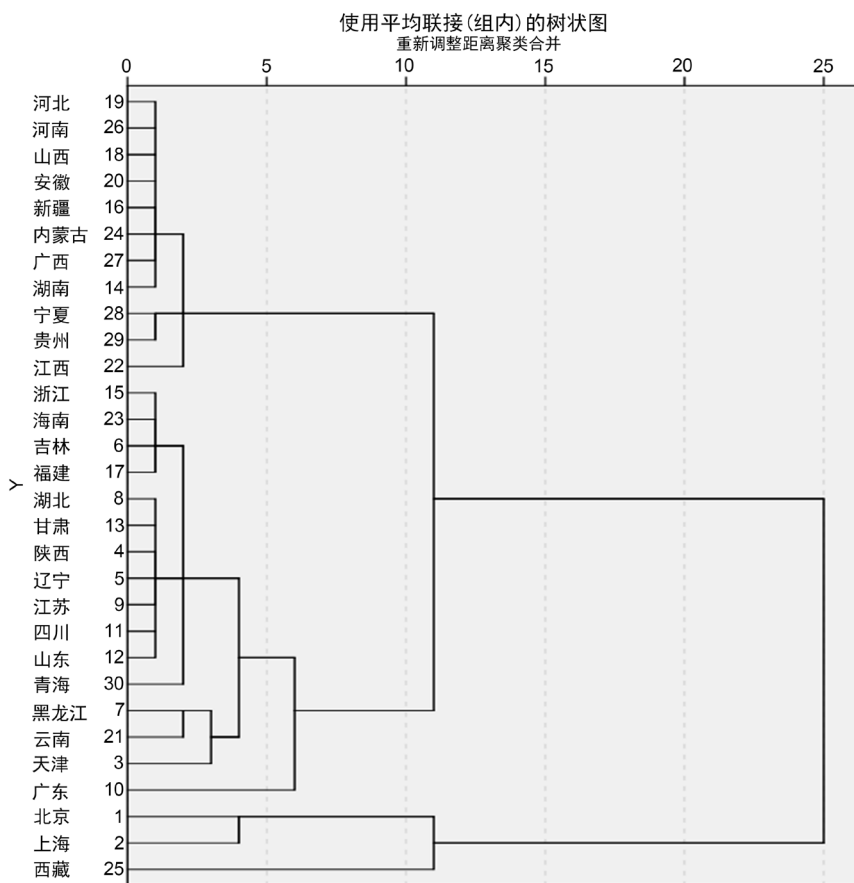
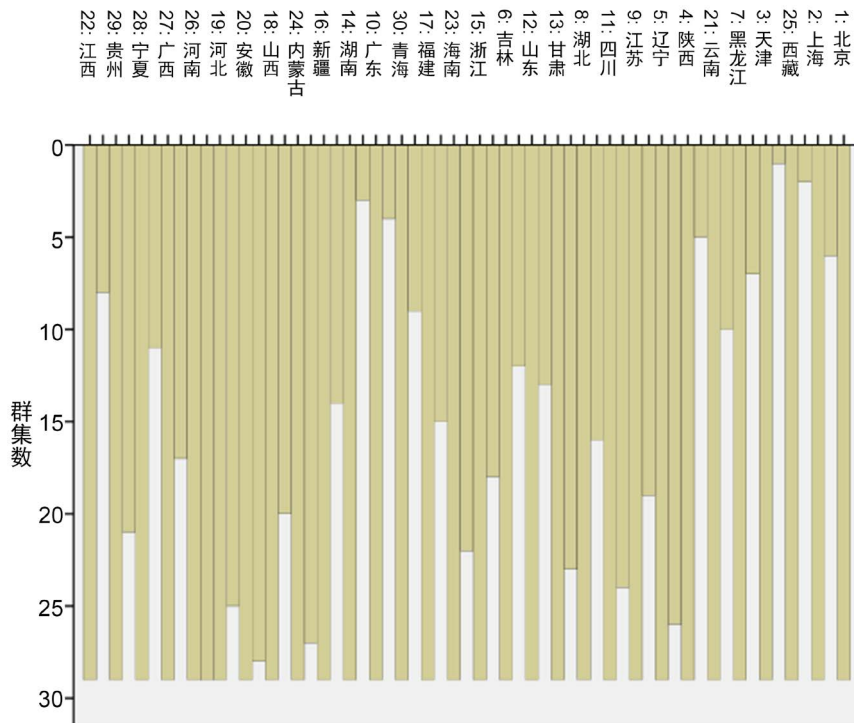


Figure 3. Clustering results of class average method

图 3. 类平均法聚类结果

2) 由图 4, 我们用 ward 方法聚成 5 类, 来分析一下高等教育发展的状况。如果用 ward 方法把我国 30 个地区分为 5 类, 结果为:

第一类: 北京、上海、广东; 第二类: 天津、黑龙江、云南; 第三类: 陕西、辽宁、吉林、湖北、江苏、四川、山东、甘肃、湖南、浙江、福建、海南、青海; 第四类: 新疆、山西、河北、安徽、江西、内蒙古、河南、广西、宁夏、贵州; 第五类: 西藏。

从以上结果还有树状图可以清晰看出, ward 方法聚类的效果的确比其他聚类方法要好, 更接近实际的情况。我们分析下, 北京、上海、广东位于教育资源最多的三个城市, 高校数目也比较多, 跟现在政治经济是吻合的, 北上广深, 人才都挤到这几个城市去, 自然教育水平发展也比较好。次于三个城市的有天津、黑龙江、云南, 这些也是比较早的文化产地, 高等教育情况也是相对比较充足。对于陕西、辽宁等城市, 这边高校资源中等, 学生数目也比较多, 但是经济发展状况不是那么良好, 教育环境有待提高, 到了后面的西藏依然是自行一类, 正是因为西藏作为一个非常特殊的少数民族地区, 并且还伴有藏独等现象, 这边的教育资源, 人口资源, 高等院校的规模都相对比较小, 导致教育资源稀缺[4]。

	群集成员	
案例		5 群集
1: 北京		1
2: 上海		1
3: 天津		2
4: 陕西		2
5: 辽宁		2
6: 吉林		2
7: 黑龙江		2
8: 湖北		2
9: 江苏		2
10: 广东		3
11: 四川		2
12: 山东		2
13: 甘肃		2
14: 湖南		4
15: 浙江		2
16: 新疆		4
17: 福建		2
18: 山西		4
19: 河北		4
20: 安徽		4
21: 云南		2
22: 江西		4
23: 海南		2
24: 内蒙古		4
25: 西藏		5
26: 河南		4
27: 广西		4
28: 宁夏		4
29: 贵州		4
30: 青海		2

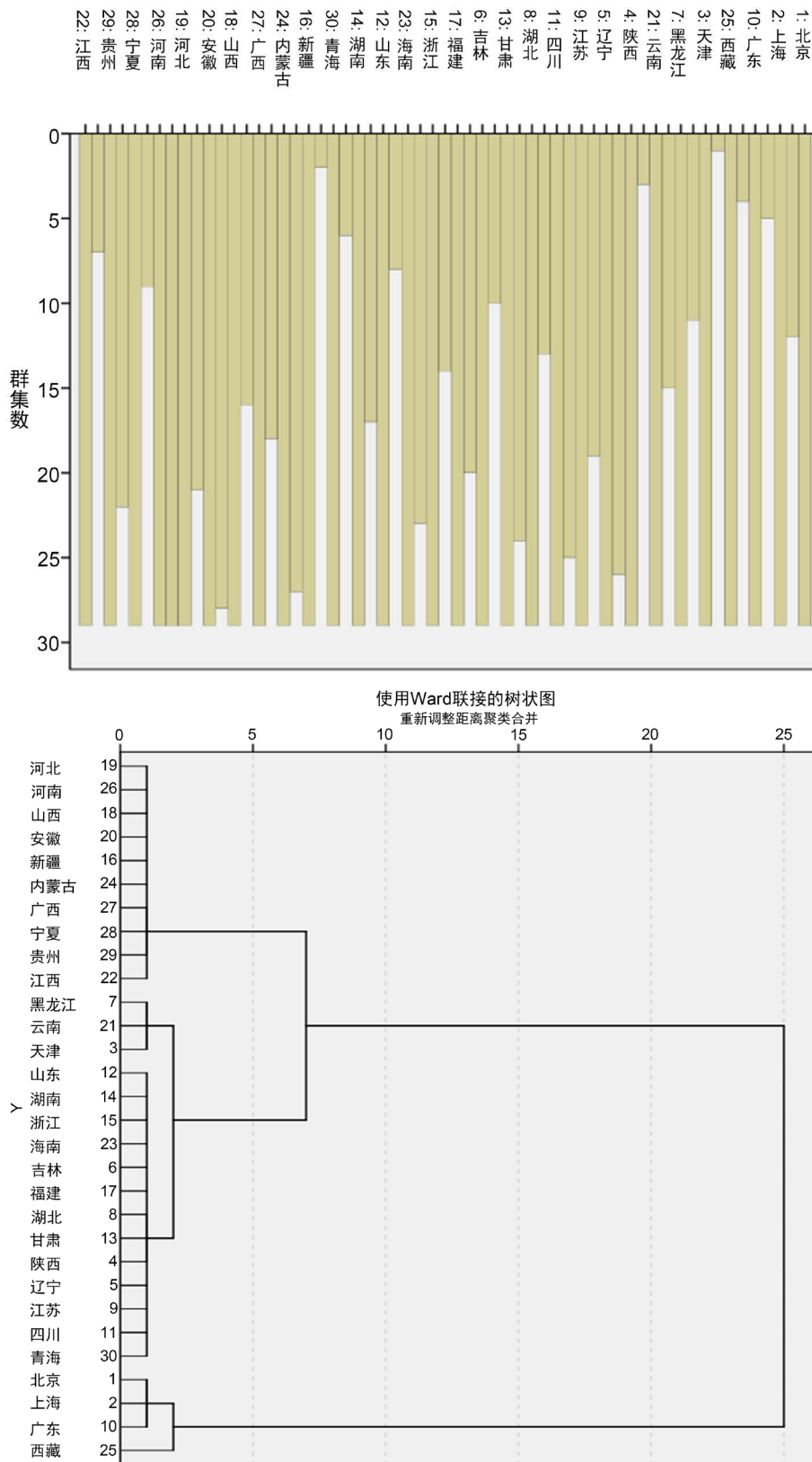


Figure 4. Ward clustering results
图 4. Ward 方法聚类结果

7. 问题与建议

根据以上的分析可以指导,在我国各省市之间,存在着严重的教育不平等的问题,无论是在硬件措施,教育水平,还是科研成果中,30个省市的教学实力都不均等。针对以上的问题我从学校、政策、社会角度提出以下建议。

1) 在偏远地方,应该多建一些学校,给每个学生提供平等教育的机会,把一些偏远地区学校的师资力量提高上来,支援外教,引进人才,制定一些福利政策,上学有优惠政策,提倡人人读书,学文化。

2) 在政策方面,国家应该加大政策力度,保障义务教育,提高教育的软实力,把一些好的硬件措施,一些珍贵的文献教材引入偏远地区的图书馆,提高学生的阅读水平。

在社会层面,应该积极号召人群给偏远山区捐款捐物,给学生提供最基本的保障,然后鼓励学生学以致用,学会了知识回到这里教书育人,给更多的学生创造更好的教育资源。提升社会公平的意识,在教育发展的过程中,特殊区域有特殊教育政策,只有全民的公平意识得到质的飞跃,才能根本上解决我国的教育不公平的问题[5]。

参考文献

- [1] 司守奎,孙兆亮. 数学建模算法与应用[M]. 第二版. 北京: 国防工业出版社, 2010.
- [2] 王学民. 应用多元统计分析[M]. 第五版. 上海: 上海财经大学出版社, 2017: 167.
- [3] 王学民. 应用多元统计分析[M]. 第五版. 上海: 上海财经大学出版社, 2017: 171.
- [4] 何嘉鹏,梁周扬,曲懿恒. 利用聚类分析研究全国教育的发展状况[J]. 中国集体经济, 2008(7): 123-124.
- [5] 晏正春. 我国教育公平的测度分析[J]. 经贸实践, 2018, 242(24): 210-211.